CAR-TR-845
CS-TR-3721

N00014-95-1-0521
December 1996

Robust Modeling and Estimation of Optical Flow with Overlapped Basis Functions

Sridhar Srinivasan* and Rama Chellappa

Center for Automation Research
and Department of Electrical Engineering
University of Maryland
College Park, MD 20742

Abstract

Computation of optical flow has been formulated as nonlinear optimization of a cost function comprising a gradient constraint term and a field smoothness factor. Results obtained using these techniques are often erroneous, highly sensitive to numerical precision, and determined sparsely, and they carry with them all the pitfalls of nonlinear optimization. In this paper, we regularize the gradient constraint equation by modeling optical flow as a linear combination of an overlapped set of basis functions. We develop a theory for estimating model parameters robustly and reliably. We prove that the extended least squares solution proposed here is unbiased and robust to small perturbations in the estimates of gradients and to mild deviations from the gradient constraint. The solution is obtained by a numerically stable sparse matrix inversion, giving a reliable flow field estimate over the entire frame. Experimental results of our scheme are surprisingly accurate and consistent across a variety of images, in comparison with the standard optical flow algorithms. We argue that our flow field model offers higher accuracy and robustness than conventional optical flow techniques, and is better suited for image stabilization, mosaicking and video compression.

Keywords: Optical Flow, Robust Techniques, Motion Analysis

The support of this research by the Defense Advanced Research Projects Agency (ARPA Order No. C635) and the Office of Naval Research under Contract N00014-95-1-0521 is gratefully acknowledged.

*LNK Corporation, Riverdale, MD 20737

1 Introduction

The estimation of inter-frame apparent motion is a key element in image sequence analysis. Accurate and dense motion estimates are a prerequisite for a wide variety of processes including image stabilization, computation of time to collision, extraction of 3D structure and video compression. The computation of *optical flow*, which is defined as the 2D projection of the 3D motion field, has been widely studied [1–12]. Recognizing that estimation of a pixelwise flow field is an ill-conditioned problem, researchers have tried to regularize it by imposing some form of smoothness. Most of the techniques can be formulated as nonlinear optimization involving a term for compliance with the *gradient constraint equation* (1) and a cost function for smoothness of the computed flow field. With this approach, the process of estimating optical flow is converted to a nonlinear optimization problem, although the starting point (1) is linear in the velocity field (u, v). Results obtained by these techniques are often significantly off the mark, highly sensitive to numerical precision, determined only at a sparse set of points with reasonable confidence, and carry with them the pitfalls associated with nonlinear optimization. On the other hand, estimation techniques that preserve the linearity of the problem by assuming a patchwise constant, affine or higher-order flow field suffer from blocking artifacts at patch boundaries. Besides, the best patch size is often determined by a poor tradeoff between localization and accuracy.

In this paper, we present a technique for modeling the optical flow field of a sequence as a linear combination of an overlapped set of basis functions. Assuming that the image sequence satisfies the gradient constraint equation, we develop a linear system for determining the model parameters. The model we choose inherently regularizes (1) and the solution to the model parameters uniquely determines the motion field. We prove that the extended least-squares solution of the system proposed here is unbiased and robust to small perturbations in the estimates of gradients and to mild, random non-compliance with the assumption (1). The linear system is sparse and its solution involves a sparse matrix inversion. A numerically stable solution is obtained by using the method of conjugate gradients, giving a dense and reliable flow field even when the matrix is ill-conditioned. We minimize boundary effects with an implementational trick that guarantees a corner-to-corner flow field. We enumerate a set of rules for choosing the basis functions, based on mathematical necessity and intuitive desirability. Finally, we compare our scheme with the standard optical flow algorithms discussed in the literature. Our results are surprisingly accurate and consistent across a wide variety of synthetic images, and agree well with our visual comprehension of real images. We conclude that the overlapped basis model for optical flow which we propose here is as accurate as, or better than, the best "generic" optical flow techniques (defined as techniques that compute the pixel-wise optical flow by non-linear optimization), despite having to conform to a model.

2 Modeling and Estimation of the Flow Field

When the projected $2D$ image field of a scene is given by $\psi = \psi(x, y, t)$, preservation of luminance patterns implies the *gradient constraint equation*

$$\frac{\partial \psi}{\partial t} + u\frac{\partial \psi}{\partial x} + v\frac{\partial \psi}{\partial y} = 0 \qquad \forall x, y, t \tag{1}$$

In (1), u and v denote the horizontal and vertical velocities (as functions of space and time), respectively. Together, they constitute the *optical flow* of the sequence. For every triplet (x, y, t) in (1), there are two unknowns, making the problem of computing the optical flow ill-conditioned. In practice, the system is regularized by imposing additional smoothness constraints on u and v. Current solutions to (1) take one of the following approaches:

- minimizing a cost function comprising a compliance term and a smoothness measure [2]
- assuming a patchwise constant, affine or polynomial model for the flow field [4,5,11]
- looking at higher order derivatives [3]
- formulating (1) as a matching problem using luminance, frequency or phase as the matching criterion [6,7,9,10].

The performance of these techniques is not consistent with respect to accuracy and reliability of the field, or to numerical sensitivity. Besides, some of the techniques are computationally expensive.

2.1 Fitting a Motion Model

An alternative to computing the optical flow on a pixelwise basis is to model the motion fields u and v in terms of a weighted sum of basis functions and estimating the weights which constitute the model parameters. In this approach, the motion field is force-fitted to the model and derives its smoothness properties from those of the model basis functions. Let $\{\phi = \phi(x, y, t)\}$ be a family of basis functions, and let the flow field be modeled as

$$u = \sum_{k=0}^{K} u_k \phi_k \quad \text{and} \quad v = \sum_{k=0}^{K} v_k \phi_k \tag{2}$$

It is not difficult to see that the patchwise constant, affine or polynomial model is a special case of (2). For example, setting ϕ_k to a rectangular window function is equivalent to the constant-in-a-patch model. Likewise, choosing a wavelet basis for $\{\phi_k\}$ is tantamount to performing a multiresolution optical flow computation. Since the optical flow field of a sequence is largely smooth, it seems reasonable to model the field using (2) and an appropriate basis $\{\phi_k\}$. What is interesting, however, is the mathematical elegance with which the model parameters $\{u_k, v_k\}$ can be extracted accurately, reliably and robustly, at a low computational cost.

Substituting (2) into (1), we get

$$\frac{\partial \psi}{\partial t} + \sum_k u_k \phi_k \frac{\partial \psi}{\partial x} + \sum_k v_k \phi_k \frac{\partial \psi}{\partial y} = 0 \quad \forall x, y, t \tag{3}$$

This system of equations in 3-space can be reduced to a scalar equation for each instant of time by integrating with a multiplicative kernel $\theta = \theta(x, y)$:

$$\int \frac{\partial \psi}{\partial t} \theta dx dy + \sum_k u_k \int \phi_k \frac{\partial \psi}{\partial x} \theta dx dy + \sum_k v_k \int \phi_k \frac{\partial \psi}{\partial y} \theta dx dy = 0 \tag{4}$$

(4) exists for every square integrable kernel θ. In order to solve for $\{u_k, v_k\}$, it is necessary to choose appropriate kernels in (4).

2.2 Solutions

System (3) is linear in the unknowns $\{u_k, v_k\}$ and is analogous to the matrix-vector system

$$Ax \rightarrow b \quad A \in \Re^{M \times N}, M > N \tag{5}$$

where x corresponds to the vector $(u_0, v_0, \ldots)'$. The analogy implies the applicability of solutions and results of (5) to (3). In the discrete domain, the analogy is obvious since an equation of type (3) exists for each pixel in the current frame, corresponding to one row of the composite matrix $[A|b]$. The least squares (LS) solution of (5) is given by $A'Ax = A'b$; choosing θ from the family $\{\phi_k \frac{\partial \psi}{\partial x}, \phi_k \frac{\partial \psi}{\partial y}\}$ gives the LS solution of (3). In practice, only discretized data is available for the image luminance field ψ. The LS solution assumes knowledge of the spatial derivatives of ψ, which may not be known reliably. Any minor and random non-compliance with (1) is accounted for in the observation error in b. A robust approach must try to minimize sensitive dependence of the solution on the spatial as well as temporal derivatives. In other words, in the analogue (5), the solution must be accurate and robust to errors in A as well as in b; this can be stated as follows:

Assume x_0 is the exact solution for the overconstrained linear system $Ax \rightarrow b$. Let Δ and δ be zero-mean, independent additive observation noise in A and b respectively, i.e. the quantities $\hat{A} = A + \Delta$ and $\hat{b} = b + \delta$ are observed. Find an "optimal" estimate x of x_0 given $\hat{A}$ and $\hat{b}$.

Our definition of "optimal" is deliberately vague. In the remainder of this section, we will analyze three solutions, *viz.* LS, total least squares (TLS), and our proposed method which we shall refer to as extended least squares (ELS), in terms of their optimality, quantified by the bias and covariance of the estimates. We will make the assumptions of uncorrelatedness of Δ and δ to A and b, and of the invertibility of the matrix $A'A$. In addition, we will assume the availability of the observation $\hat{A}'A$ for the ELS estimate, the existence of which we shall prove later. The symbol $\approx$ denotes a first-order approximation. R_Δ denotes the covariance matrix of the rows of Δ and r_δ is the variance of δ_i, assuming that the rows are *iid.*

2.2.1 LS solution

The LS solution of (5) is $x_{LS} = (\hat{A}\hat{A}')^{-1}\hat{A}'b$ with an error $e = x - x_0$ given by

$$e_{LS} \approx (A'A)^{-1}A'(\delta - \Delta x_0) - \underbrace{(A'A)^{-1}A'R_\Delta x}_{\text{bias}} \tag{6}$$

x_{LS} is the minimizer of $\|Ax - b\|_2$ when the observation error is present only in $\hat{b}$. Bias occurs when $\hat{A} \neq A$. The two random components of e are independent by the assumption of independence of Δ and δ. A modification of LS, the Corrected Least Squares (CLS) method, removes bias while preserving the error covariance. The CLS estimate, which is given by

$$x_{CLS} = (\hat{A}\hat{A}' - R_\Delta)^{-1}\hat{A}'\hat{b} \tag{7}$$

assumes that R_Δ is known.

2.2.2 TLS solution

The TLS principle has emerged as an alternative to LS since it is capable of handling errors in potentially all the observations in a linear system, not merely erroes on the "right hand side" [13]. The TLS solution is obtained by minimally perturbing the composite observation matrix $[\hat{A}|\hat{b}]$ to reduce its rank to N. x_{TLS}, the TLS estimate, is given by

$$x_{TLS} = -\frac{1}{v_{n+1,n+1}} v_{n+1}^{(n)} \tag{8}$$

where $v_{n+1}^{(n)}$ is the vector formed by the first n components of v_{n+1}, which is the eigenvector of $[\hat{A}|\hat{b}]'[\hat{A}|\hat{b}]$ corresponding to its smallest eigenvalue μ. $v_{n+1,n+1}$ is the $n+1$th component of v_{n+1}.

e_{TLS} is not easy to determine for the TLS case, even when the observation errors are small. The coupled expressions for e_{TLS} and μ, to a first order approximation, are given by

$$\begin{aligned} \mu &\approx r_\delta + 2b'\delta - (\delta'A + b'\Delta)x_0 - b'Ae_{TLS} \\ \mu x_0 &\approx (A'\Delta + \Delta'A + R_\Delta)x_0 + A'Ae_{TLS} - \Delta'b - A'\delta \end{aligned} \tag{9}$$

It is cumbersome to analyze the bias and covariance of the TLS solution. Setting $E[e_{TLS}] = 0$ in (9) gives as necessary condition for a zero bias estimate

$$(r_\delta I - R_\Delta)x_0 \to 0 \tag{10}$$

The significance of (10) to (3) is that the TLS estimate is unbiased *only if* the error in estimating the temporal gradient is equal in variance to the error in estimating the (windowed) spatial gradient. However, in typical image sequence problems, the temporal gradient is calculated over a smaller number of frames than the spread of the spatial gradient operator. In order to satisfy (10), it is necessary to discard useful information by narrowing down the support of the spatial gradient operator, which is not desirable. In addition, it has been argued in [13] that the covariance of an unbiased TLS estimate is larger than that of the LS estimate, in the first order approximation as well as in simulations. In effect, therefore, there is no fundamental gain in choosing the TLS over the LS solution.

2.2.3 ELS solution

Neither the LS nor the TLS solution of (5) is unbiased in the general case, and the CLS solution (7) shows sensitive dependence on errors in $\hat{A}$ as well as in $\hat{b}$. Is it possible to do better? The answer is in the affirmative if an additional observation, *viz.* $\hat{G} = \hat{A}'A$, is available. What is surprising is that (i) *given the observations* $\hat{G}$ *and* $\hat{A}'\hat{b}$, *the ELS solution proposed here shows no dependence on the error* Δ *in the estimate of* A, and (ii) *the observation corresponding to* $\hat{G}$ *is available for* (3). We will prove claim (i) in this section and deal with (ii) in the next.

The ELS solution of (5) is the solution of $\hat{G}x = \hat{A}\hat{b}$,

$$x_{ELS} = \hat{G}^{-1}\hat{A}\hat{b} \tag{11}$$

which has the estimation error

$$\begin{aligned} e_{ELS} &= \underbrace{\hat{G}^{-1}}_{\approx(A'A)^{-1}} [\Delta' \underbrace{(Ax_0 - b)}_{=0} + \underbrace{\Delta'\delta}_{O(2)} + A'\delta] \\ &\approx (A'A)^{-1}A'\delta \end{aligned} \tag{12}$$

proving claim (i). x_{ELS} is unbiased, and has a smaller covariance than x_{LS} or x_{CLS} since it does not have the "Δ" term of the non-bias error in (6). In the original problem (3), the ELS solution is obtained when θ in (4) is chosen from the family $\{\phi_k \frac{\partial\hat{\psi}}{\partial x}, \phi_k \frac{\partial\hat{\psi}}{\partial y}\}$ where the quantity $\frac{\partial\hat{\psi}}{\partial\{x,y\}}$ is an *estimate* of the derivative, giving the system

$$\begin{aligned} \int \frac{\partial\hat{\psi}}{\partial t}\phi_l \frac{\partial\hat{\psi}}{\partial x}dxdy + \sum_k u_k \int \phi_k \frac{\partial\psi}{\partial x}\phi_l \frac{\partial\hat{\psi}}{\partial x}dxdy + \sum_k v_k \int \phi_k \frac{\partial\psi}{\partial y}\phi_l \frac{\partial\hat{\psi}}{\partial x}dxdy &= 0 \\ \int \frac{\partial\hat{\psi}}{\partial t}\phi_l \frac{\partial\hat{\psi}}{\partial y}dxdy + \sum_k u_k \int \phi_k \frac{\partial\psi}{\partial x}\phi_l \frac{\partial\hat{\psi}}{\partial y}dxdy + \sum_k v_k \int \phi_k \frac{\partial\psi}{\partial y}\phi_l \frac{\partial\hat{\psi}}{\partial y}dxdy &= 0 \end{aligned} \tag{13}$$

with the estimated temporal derivative $\frac{\partial\hat{\psi}}{\partial t}$. We will now prove the availability of the observations $\hat{G} = \hat{A}'A$ and $\hat{A}'\hat{b}$, which is equivalent to proving the computability of the integrals in (13), under certain weak assumptions on the functional form of ϕ_k and the estimate $\frac{\partial\hat{\psi}}{\partial\{x,y\}}$.

2.3 Eliminating Derivatives

Consider the integral $I(y) = \int \phi_k \frac{\partial\psi}{\partial x}\phi_l \frac{\partial\hat{\psi}}{\partial x}dx$. Assume that the estimate $\frac{\partial\hat{\psi}}{\partial\{x,y\}}$ has a differentiable functional form, *i.e.* the derivative $\frac{\partial}{\partial\{x,y\}}\frac{\partial\hat{\psi}}{\partial\{x,y\}}$ is known exactly. This will hold for even the simplest of discrete gradient masks like $(\ldots, 0, -1, 1, 0, \ldots)$ since the masks assume a smooth underlying functional form. Also, assume that the ϕ_k are differentiable and that $\phi_k(x, y) \to 0$ as $x \to \pm\infty$ or $y \to \pm\infty$. Integrating $I(y)$ by parts over $(-\infty, \infty)$, we get

$$I(y) = \underbrace{\left[\phi_k\phi_l \frac{\partial\hat{\psi}}{\partial x}\psi\right]_{-\infty}^{\infty}}_{=0} - \int_{-\infty}^{\infty} \psi \frac{\partial\phi_k\phi_l\frac{\partial\hat{\psi}}{\partial x}}{\partial x}dx \tag{14}$$

which is computable reliably without knowing the exact derivatives $\frac{\partial\psi}{\partial\{x,y\}}$. Applying this reasoning to (13) gives

$$\begin{aligned}
\sum_k u_k \int \frac{\partial \phi_k \phi_l \frac{\partial\hat{\psi}}{\partial x}}{\partial x}\psi \; + \; \sum_k v_k \int \frac{\partial \phi_k \phi_l \frac{\partial\hat{\psi}}{\partial x}}{\partial y}\psi \;\; &= \;\; \int \frac{\partial\hat{\psi}}{\partial t}\phi_l \frac{\partial\hat{\psi}}{\partial x} \\
\sum_k u_k \int \frac{\partial \phi_k \phi_l \frac{\partial\hat{\psi}}{\partial y}}{\partial x}\psi \; + \; \sum_k v_k \int \frac{\partial \phi_k \phi_l \frac{\partial\hat{\psi}}{\partial y}}{\partial y}\psi \;\; &= \;\; \int \frac{\partial\hat{\psi}}{\partial t}\phi_l \frac{\partial\hat{\psi}}{\partial y}
\end{aligned} \tag{15}$$

where the integrals are over the entire X-Y plane and can be computed reliably. (15) has the following desirable properties:

- The accuracy of the spatio-temporal image derivatives is not critical to the accuracy of computation.
- The computed image flow is force-fitted to a model. The only conditions on the model are that it be space-limited and differentiable.
- With finite-extent basis functions ϕ_k, the system of equations has a sparse, banded matrix structure.

2.4 Choice of Basis Functions

As in all estimation problems, there is the inevitable tradeoff between sensitivity and selectivity, or equivalently, between accuracy and localization. The specific basis function family $\{\phi_k\}$ determines the shape of the ROC curve, and its density (defined as the number of basis functions per unit area) fixes the operating point on the curve. Not all differentiable, compactly supported curves are meaningful to the problem at hand. Since the variations in optical flow are typically high-frequency and localized or low-frequency and global, periodic bases do not offer any advantage. A point to note here is that orthogonal bases do not lead to a simpler solution than non-orthogonal bases since there are no integrals of the form $\int \phi_k \phi_l$. However, for every non-overlapping pair $\{\phi_i, \phi_j\}$, the corresponding entries in the observation matrix $\hat{G}$, $\hat{G}_{ij}$ and $\hat{G}_{ji}$, are zero. Thus, from the computational aspect, it is desirable to minimize the number of overlapping pairs.

The reasoning thus far, including the final linear system (15), holds for any choice of basis functions which are differentiable and which decay to zero. In the remainder of this section, we will place certain additional restrictions on the choice of $\{\phi_k\}$ that ensure computational ease and have intuitive appeal for modeling a motion field. In our experiments, we constructed $\{\phi_k\}$ from translations of a prototype function ϕ_0 along a uniformly spaced square grid of spacing w. The additional requirements we place on ϕ_0 are the following:

- *Separability:* $\phi_0(x,y) = \phi_0(x)\phi_0(y)$
- *Symmetry about the origin:* $\phi_0(x) = \phi_0(-x)$
- *Peak at the origin:* $|\phi_0(x)| \leq \phi_0(0) = 1$

- *Compact support:* $\phi_0(x) = 0 \quad \forall |x| > w$
- *Constancy:* $\phi_0(x) + \phi_0(x - w) \rightarrow 1, \forall x \in [0, w]$
- *Linearity:* $\phi_0(x) + \lambda\phi_0(x - w) \rightarrow \lambda + (\lambda - 1)\frac{x}{w}\ \forall x \in [0, w]$

Note that *constancy* is a special case of *linearity*. *Compact support* ensures that each basis function overlaps with exactly eight of its neighbors, in the cardinal and diagonal directions. *Constancy* is essential for modeling the simple case of uniform translation, and *linearity* is necessary for exactly modeling an affine flow field. Even if the requirement of *compact support* were to be waived in view of the rapid decay of the Gaussian function, it still does not satisfy *linearity* or even *constancy* unless the definition is broadened to cover an arbitrarily large linear combination. In our experiments we have used the cosine window

$$\phi_{c,0}(x) = \frac{1}{2}\left[1 + \cos(\frac{\pi x}{w})\right] \quad x \in [-w, w] \tag{16}$$

This function satisfies *linearity* only approximately, but has continuous derivatives and performs very well. The bilinear interpolator

$$\phi_{b,0}(x) = \left(1 - \frac{|x|}{w}\right)^+ \tag{17}$$

satisfies all requirements and performs satisfactorily although not as well as the cosine window. We believe that this is due to second-order effects of the discontinuity in its derivative.

The spacing w of the grid at which the basis functions are centered determines the density of the basis functions, thereby determining the position of the algorithm on the ROC curve. A larger spacing implies a larger support for ϕ_k which reduces error in the estimates of the integrals in (15). However, increasing w means that fewer basis functions (or degrees of freedom) are available for modeling the flow field. Trying to fit a stiff model can potentially lead to non-compliance. w, therefore, determines the tradeoff point between robustness and accuracy of the algorithm. In our experiments, we used the rule

$$w = \min\{32, \frac{r}{5}, \frac{c}{5}\} \tag{18}$$

where the image size was $r \times c$.

2.5 Boundary Effects

Since spatial gradients are typically computed by convolving the image with a finite impulse response filter, gradient estimates are absent at the periphery of the image within a margin determined by the length of the operator. Generic optical flow algorithms often do not estimate the flow within this margin due to non-availability of the gradients. Flow estimates along the frame border are necessary in mosaicking and video compression. In particular, since the emerging area in a panned sequence of images is at the periphery, a motion estimate at the periphery is necessary in order to stitch together the image

sequence into one panoramic view. Instead of having no boundary flow estimate at all, an estimate with a graceful degradation in its reliability, in consonance with the reduced information available, is a preferable alternative.

Within a margin near the periphery, the modeled optical flow is determined by the coefficients corresponding to basis functions both inside and outside the image frame. By enforcing spatial continuity on the coefficients, it is possible to reduce the dependence to coefficients within the image frame. We have found zero-order continuity to be satisfactory, although in theory first-order continuity is necessary for modeling affine flow. Zero-order continuity is enforced by replacing the out-of-frame coefficients (u_k, v_k) by the nearest peripheral coefficients (u_l, v_l) in the equations corresponding to peripheral ϕ_ls. Integration is performed over the sub-region for which a gradient estimate is available. Since the domain of integration is not infinite, (14) no longer holds, and (13) must necessarily be used for peripheral coefficients. Using derivatives leads to a loss in robustness at the periphery, which is also affected by the integration window being smaller than the support of ϕ_l. This is the cost incurred for obtaining a corner-to-corner flow field estimate.

2.6 Sparse Matrix Inversion

If the image frame is covered by M basis functions of which M_p are peripheral, we have $2M_p$ equations of the form (13) and $2(M - M_p)$ equations of the form (15). The integrals on the LHS are aggregated into the observation matrix $\hat{G}$ which is ideally symmetric and positive semi-definite. M is determined from the grid spacing w and the image size. The solution for the coefficients $\{u_k, v_k\}$ involves the inversion of $\hat{G}$, which is a block diagonal dominant $2M \times 2M$ matrix, with only $O(M)$ nonzero entries, when ϕ_0 satisfies *compact support.* We have used the conjugate gradient method [14] to perform this inversion. This method takes $O(M)$ iterations for convergence, each iteration involving three sparse matrix multiplications of complexity $O(M)$, leading to an overall complexity of $O(M^2)$. In theory, the method will converge to the correct solution if the matrix is symmetric and positive definite. In practice, we have never encountered a divergent situation for a wide variety of imagery and severely ill-conditioned $\hat{G}$. Since $\hat{G}$ is block diagonally dominant in 2×2 blocks, a good starting point is the solution to the 2×2 diagonal subsystem for each pair (u_k, v_k). In other words, the initial guess is the solution to (13) or (15) with the summand set to zero when $k \neq l$. When the submatrices are singular, the corresponding starting values are set to zero. In the event that $\hat{G}$ is singular, the implementation in [14] gives the best least-squares approximation to the system. This is reasonable since singularity implies that there exist substantial areas in the image with zero or strictly one-dimensional spatial gradient, in which case the optical flow is ill-defined. For small regions with zero gradient, continuity of the direction of flow is effected by contributions to the flow from neighboring basis functions.

3 Results

In keeping with the practice established in [1], we performed a series of experiments on synthetic and real data. Exact flows are known for the synthetic data, which consist of

Experiment	Sinusoid		Square	
	$\bar{\epsilon}$	$\bar{\sigma}$	$\bar{\epsilon}$	$\bar{\sigma}$
Horn & Schunck (original)	4.19	0.50	47.21	14.60
Horn & Schunck (mod.)	2.55	0.59	32.81	13.67
Lucas & Kanade	2.47	0.16	×	×
Uras *et al.*	2.59	0.71	×	×
Nagel	2.55	0.93	34.57	14.38
Anandan	30.80	5.45	31.46	18.31
Singh ($n = w = 2, N = 2$)	2.24	0.02	49.03	21.38
Singh ($n = w = 2, N = 4$)	91.71	0.04	45.16	21.10
Fleet & Jepson ($\tau = 1.25$)	0.03	0.01	×	×
Proposed	2.46	0.03	23.95	24.36

Table 1: Performance on *Sinusoid* and *Square* data

the *Sinusoid, Square, Translating tree, Diverging tree* and *Yosemite* sequences (Figs. 1 and 2). *Sinusoid* is a uniformly translating modulated 2D pattern with a wavelength of 6 pixels. *Square* is the simple case of a white square moving against a dark background. The *tree* sequences are synthesized from realistic data simulating camera translation with respect to the 2D scene. *Yosemite* is a 2D rendering of a 3D model, with motion discontinuities introduced at the top of the mountain range. The clouds in this sequence translate uniformly across the image while undergoing a steady luminance change. In addition, there is significant aliasing near the lower portions of the image, making this sequence particularly challenging. The *SRI trees, Coke can, Rubik cube* and *Taxi* sequences comprise the real data (Figs.3-5).

An error measure can be defined for estimates of the optical flow for synthetic data sequences, since exact flows are known. We use the angular error measure employed in [1]. Assume that the true and computed flows at a point (x, y) in a particular frame are $(u_0, v_0)'$ and $(u, v)'$ respectively. Define vectors $\mathbf{v}_0 = (u_0, v_0, 1)'$ and $\mathbf{v} = (u, v, 1)'$. The error angle ϵ at (x, y) is given by

$$\epsilon = \arccos\left(\frac{\mathbf{v}_0.\mathbf{v}}{\|\mathbf{v}_0\|\|\mathbf{v}\|}\right) \tag{19}$$

ϵ is insensitive to the magnitude of the motion vector and offers a normalized measure against which a range of velocities can be compared meaningfully. The mean value of ϵ over all image points where an optical flow estimate is available is the final scalar error metric $\bar{\epsilon}$. The standard deviation $\bar{\sigma}$ of ϵ indicates the spread of errors over the image frame. $\bar{\epsilon}$ and $\bar{\sigma}$, for the generic optical flow algorithms, have been obtained from [1]. For the real sequences, exact flows are not known and the performance is evaluated subjectively by visual analysis. The flows shown in Figures 1-5 are suitably downsampled and rescaled. The tabulated values of $\bar{\epsilon}$ and $\bar{\sigma}$ are in degrees.

In order to keep the comparison equitable, we have only considered optical flow techniques giving a 100% dense flow field in Tables 1 and 2. Even among these, many algorithms estimate optical flow only over an inscribed region whose margin is determined by the spread of the gradient operator, while the proposed technique gives a corner-to-corner flow estimate. For the *Sinusoid* data set (Fig. 1(a-b), Table 1), the proposed technique

Experiment	Translating Tree		Diverging Tree		Yosemite	
	$\bar{\epsilon}$	$\bar{\sigma}$	$\bar{\epsilon}$	$\bar{\sigma}$	$\bar{\epsilon}$	$\bar{\sigma}$
Horn and Schunck (original)	38.72	27.67	12.02	11.72	31.69	31.18
Horn and Schunck (modified)	2.02	2.27	2.55	3.67	9.78	16.19
Uras *et al.* (unthresholded)	0.62	0.52	4.64	3.48	8.94	15.61
Nagel	2.44	3.06	2.94	3.23	10.22	16.51
Anandan	4.54	3.10	7.64	4.96	13.36	15.64
Singh (step 1, $n = 2$, $w = 2$)	1.64	2.44	17.66	14.25	15.28	19.61
Singh (step 2, $n = 2$, $w = 2$)	1.25	3.29	8.60	5.60	10.44	13.94
Proposed	0.61	0.26	2.94	1.64	8.94	10.63

Table 2: Performance on *Translating tree*, *Diverging tree* and *Yosemite*

ranks after Fleet and Jepson's, which benefits greatly from the sharp frequency spectrum of the input sinusoid, and Singh's method, which is comparable in accuracy. The proposed technique is the best for *Square* (Fig. 1(c-f), Table 1) by a wide margin. For the *Square* sequence, Fleet and Jepson's algorithm is unable to produce a 100% dense flow field, and $\bar{\epsilon}$ of Singh's algorithm is greater than twice $\bar{\epsilon}$ of the proposed algorithm. Interestingly, while the "true" flow field of *Square* has been deemed to be a uniform translation across the image (Fig. 1(d)), it is a moot point whether the background is indeed moving along with the white square. The output of our algorithm (Fig. 1(e)) shows only the square in motion, which is an interpretation as credible as the "true" flow. On the other hand, Anandan's algorithm (Fig. 1(f)), which ranks second in Table 1, computes flow only for approximately half the image area due to border effects, and the computed flow pattern is visually meaningless. The proposed technique performs very well for the synthetic data sets *Diverging tree, Translating tree* and *Yosemite* (Table 2). The results are shown in Fig. 2 where the three columns correspond to the center frame, true flow and computed flow respectively. Our algorithm ranks at the top for *Translating tree* and *Yosemite* and in second place for *Diverging tree.* Note that the error standard deviation $\bar{\sigma}$ is significantly smaller than the $\bar{\sigma}$ of the other techniques. The results indicate that the proposed technique is consistently accurate over the range of synthetic imagery.

Likewise, for the real data sets (Figs. 3-5) we can see that the computed flows are dense and visually consistent with the observed motion. In contrast with the results in [1], it can be seen that our flows do not show large random errors, and all reasonably-sized objects in motion are picked up. The computed flow is quite accurate—for example, the rotational flow at the top of the Rubik cube shows up in detail. For the *Taxi* sequence, all vehicles in motion are captured, and for *SRI trees* the direction of translation is computed accurately. The divergent motion of the *Coke can* sequence shows up clearly despite the lack of detail and presence of occluding edges. For a panning motion, it is possible to compute a pseudo-depth map on the basis of the flow field. Such a map is shown in Fig. 5(c) for *SRI trees*, using the computed flow (Fig. 5(b)). The ground and branches close to the camera show up as bright regions while the distant background shows up as dark. The limitation of using a model-based approach is in its inability to capture highly localized motion as seen in the non-appearance of the pedestrian in the top left area of *Taxi* on the flows. However, an examination of the flows in [1] reveals that *none* of the

generic methods succeed in locating the human, and many even fail to locate the two dark vehicles in motion. The only inaccuracy we can find from a visual inspection of the flows is a small kink near the fork of the tree in *SRI trees* which is due to occlusion of the background, leading to a structured non-compliance with (1).

From our experiments, we can infer that the primary source of error in our technique is non-compliance with (1), most often as result of occlusion or change in illumination. It is possible to introduce a luminance term in (1) and perform a similar analysis. Likewise, it is also possible to integrate information across time either by assuming a constant motion model or by extending the domain of the basis function set $\{\phi_k\}$ over the temporal dimension. We conjecture that either of these extensions will improve the accuracy of the flow estimates at the cost of increased computation.

4 Conclusion

The results of the proposed technique clearly indicate its superiority relative to generic optical flow methods in accuracy, robustness and reliability for a range of synthetic and real data. Apart from the primary requirements of accuracy, smoothness and computational ease, there are secondary, application-specific requirements that determine which algorithm is used for motion analysis in a computer vision system. Situations exist where a model-based optical flow estimate is preferred over a generic optical flow estimate. For example, in 2D image stabilization, the flow field is consolidated into a small set of (similarity or affine) parameters that characterize global interframe motion. Outliers like foreground objects can severely impair mean-square estimates of global motion from the pixelwise flow field. When the optical flow is represented in terms of a small number of parameters, it is possible to apply a robust but computationally intensive estimator based on, for instance, the Hough transform to compute global motion parameters. This technique becomes increasingly infeasible in the generic optical flow case due to the increased data size.

Similarly, in the context of video compression, a flow field characterized by a model is preferable to a generic flow field. Encoding a generic flow field for transmission over a channel or for storage involves increased complexity, higher bandwidth, and more errors due to quantization than does the process of encoding a set of model parameters. The only benefit in having a generic description of the field is that motion discontinuity information, if it is extracted reliably, can aid the compression process. However, for video compression, it can be argued that the cost associated with encoding motion discontinuities is sufficiently high to negate the advantage of using a generic flow field. Thus, the loss of generality in requiring that the flow field conform to a model is no less advantageous. In light of the above, we conclude that the robust optical flow modeling and estimation technique proposed here is better in performance and wider in applicability than the standard optical flow methods in the literature.

References

[1] J. L. Barron, D. J. Fleet and S. S. Beauchemin, "Performance of Optical Flow Techniques", *Int. Jour. Comp. Vision*, vol. 12, pp. 43-77, 1994. Sequences and code available at ftp://ftp.csd.uwo.ca/pub/vision.

[2] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow", *Artificial Intelligence*, vol. 17, pp. 185-204, 1981.

[3] H. H. Nagel, "On the Estimation of Optical Flow", *Artificial Intelligence*, vol. 33, pp. 299-324, 1987.

[4] S. Uras *et. al.*, "A Computational Approach to Motion Perception", *Biological Cybernetics*, vol. 60, pp. 79-97, 1988.

[5] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", *Proc. DARPA IUW*, pp. 121-130, 1981.

[6] P. Anandan, "Measuring Visual Motion from Image Sequences", Ph. D. dissertation, University of Massachusetts, Amherst, 1987.

[7] A. Singh, "An Estimation-Theoretic Framework for Image-Flow Computation", *Proc. IEEE ICCV*, pp. 168-177, 1990.

[8] A. M. Waxman, J. Wu and F. Bergholm, "Convected Activation Profiles and the Measurement of Visual Motion", *Proc. IEEE CVPR*, pp. 717-723, 1988.

[9] D. J. Heeger, "Model for the Extraction of Image Flow", *Jour. Opt. Soc. Amer.*, vol. 4, pp. 1455-1471, 1987.

[10] D. J. Fleet and A. D. Jepson, "Computation of Component Image Velocity from Local Phase Information", *Int. Jour. Comp. Vision*, vol. 5, pp. 77-104, 1990.

[11] N. Gupta, "Recovering Shape and Motion from a Sequence of Images", Ph. D. dissertation, University of Maryland, College Park, 1993.

[12] H. Liu, "A General Motion Model and Spatio-Temporal Filters for 3-D Motion Interpretations", Ph. D. dissertation, University of Maryland, College Park, 1995.

[13] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem — Computational Aspects and Analysis*, SIAM, 1991.

[14] W. H. Press *et. al.*, *Numerical Recipes in C*, Cambridge University Press, 1989.

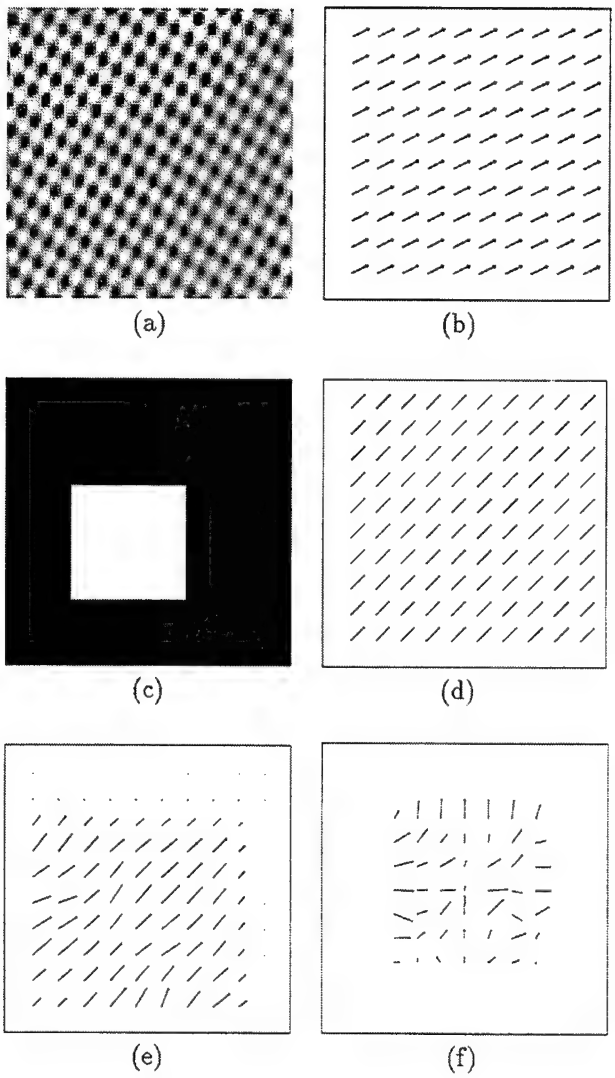


Figure 1: *Sinusoid*: (a) center frame, (b) computed flow (visually identical to true flow); *Square*: (c) center frame (d) true flow (e) computed flow using proposed algorithm (f) computed flow by Anandan's algorithm.

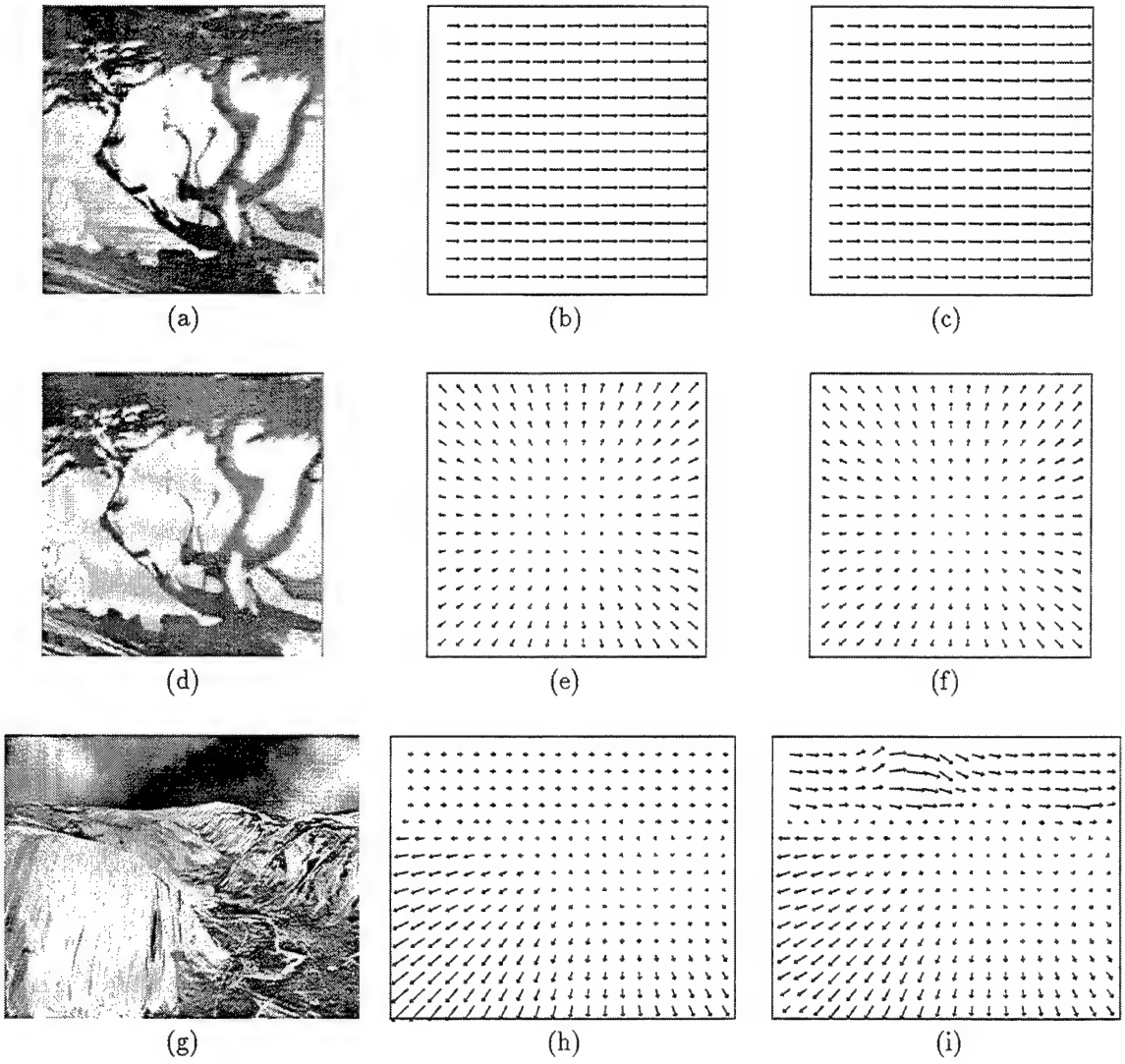


Figure 2: (a)-(c) *Translating tree*, (d)-(f) *Diverging tree*, and (g)-(i) *Yosemite* sequences; (a),(d),(g) center frames, (b),(e),(h) true flows, and (c),(f),(i) computed flows.

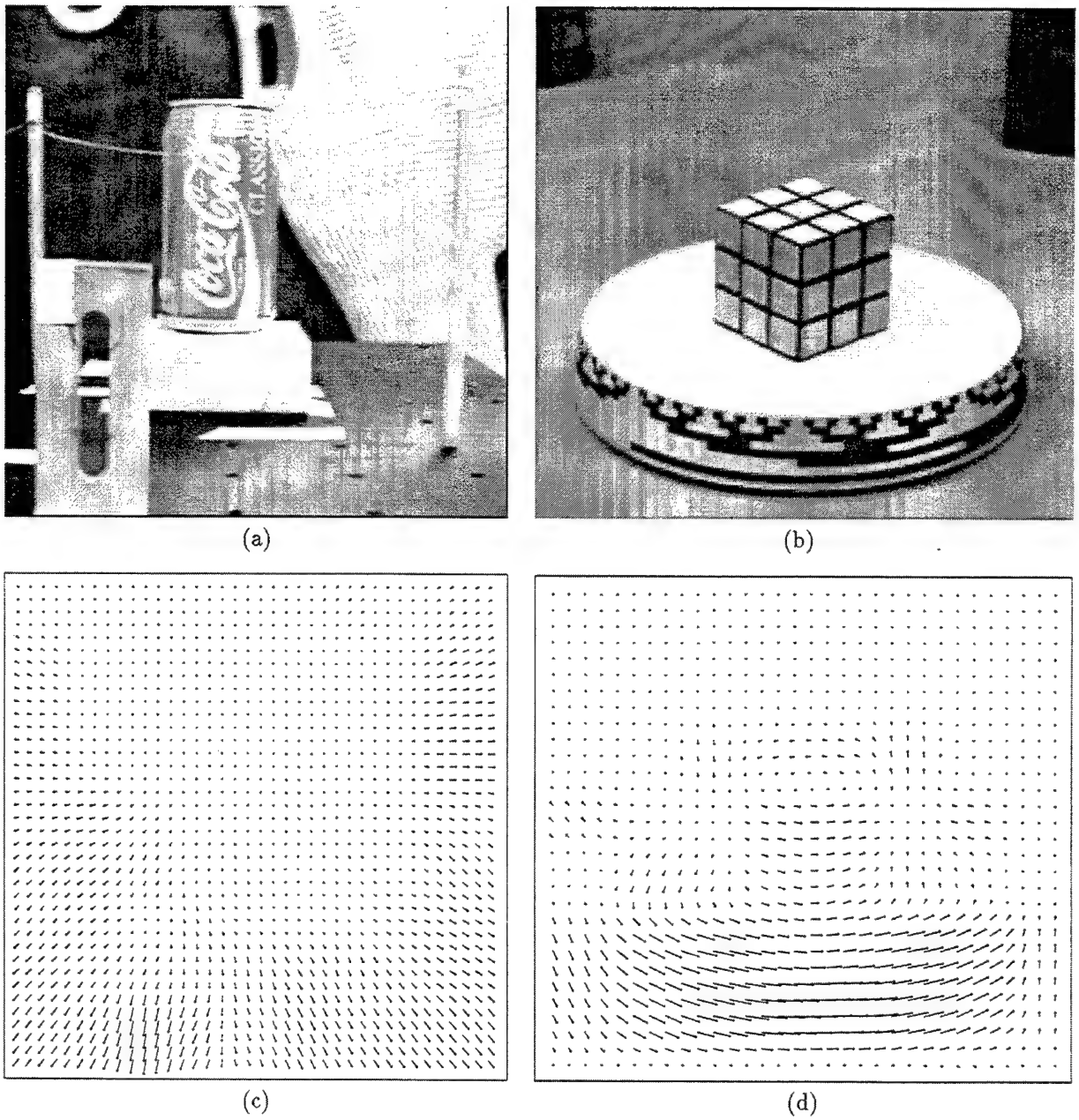
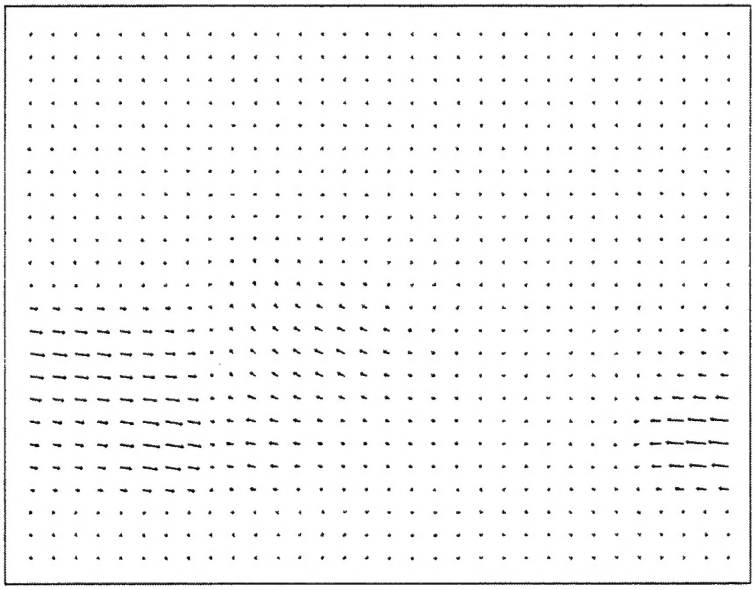


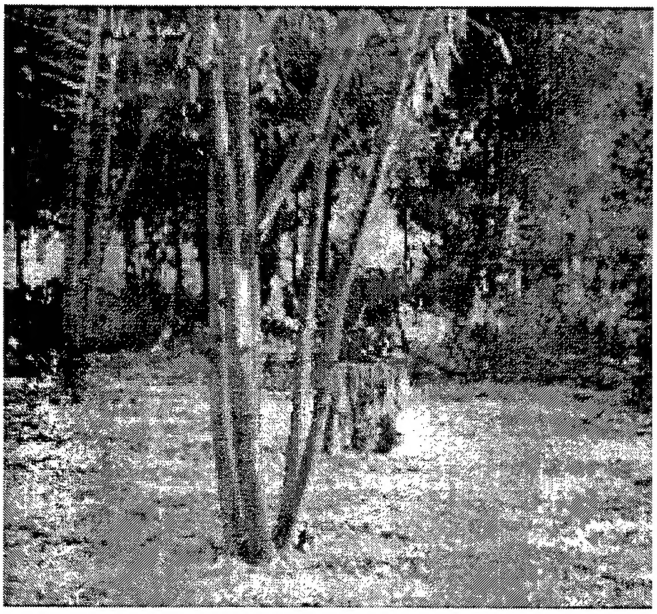
Figure 3: (a),(b) *Coke can* and *Rubik* sequence center frames; (c),(d) computed flows.

(a)

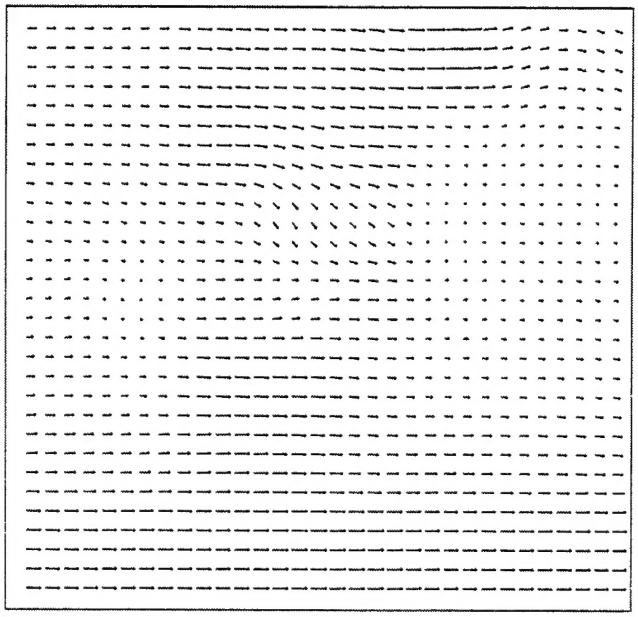


(b)

Figure 4: (a) *Taxi* center frame; (b) computed flow.



(a)



(b)

(c)

Figure 5: *SRI trees* (c) center frame, (d) computed flow, and (e) pseudo-depth map.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1996	Technical Report

4. TITLE AND SUBTITLE

Robust Modeling and Estimation of Optical Flow with Overlapped Basis Functions

5. FUNDING NUMBERS

N00014-95-1-0521

6. AUTHOR(S)

Sridhar Srinivasan and Rama Chellappa

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Automation Research
and Department of Electrical Engineering
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

CAR-TR-845
CS-TR-3721

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 North Quincy Street, Arlington, VA 22217-5660

Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Computation of optical flow has been formulated as nonlinear optimization of a cost function comprising a gradient constraint term and a field smoothness factor. Results obtained using these techniques are often erroneous, highly sensitive to numerical precision, and determined sparsely, and they carry with them all the pitfalls of nonlinear optimization. In this paper, we regularize the gradient constraint equation by modeling optical flow as a linear combination of an overlapped set of basis functions. We develop a theory for estimating model parameters robustly and reliably. We prove that the extended least squares solution proposed here is unbiased and robust to small perturbations in the estimates of gradients and to mild deviations from the gradient constraint. The solution is obtained by a numerically stable sparse matrix inversion, giving a reliable flow field estimate over the entire frame. Experimental results of our scheme are surprisingly accurate and consistent across a variety of images, in comparison with the standard optical flow algorithms. We argue that our flow field model offers higher accuracy and robustness than conventional optical flow techniques, and is better suited for image stabilization, mosaicking and video compression.

14. SUBJECT TERMS

Optical Flow, Robust Techniques, Motion Analysis

15. NUMBER OF PAGES

20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- DOE - See authorities.
- NASA - See Handbook NHB 2200.2.
- NTIS - Leave blank.

Block 12b. Distribution Code.

- DOD - Leave blank.
- DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- NASA - Leave blank.
- NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.